

STATISTICS

(25-30 marks) (P1+P2)

(EASIEST PORTION)

Bad $\rightarrow$ A*

Bad $\rightarrow$ D/E

Personal Deadline (10 Nov)

1) Notes Yaad

2) P1 worksheet.

3) P1 worksheet twice

- ✓ 1) Mode - Median - Mean - Range
- ✓ 2) Histograms
- ✓ 3) Cumulative Frequency
- ✓ 4) Correlation, Scatter Plot.

MODE - MEDIAN - MEAN - RANGE

RAW DATA (NO FREQUENCY TABLE)

10	8	7	8	10	6	11	10
10	5	7	6	8	9	10	10

Total = n = 16

1) MODE: (MOST OCCURRING VALUE) = 10

2) MEDIAN: MIDDLE VALUE

STEP 1: ARRANGE IN INCREASING ORDER

5, 6, 6, 7, 7, 8, 8, 8, 9, 10, 10, 10, 10, 10, 10, 11

X X X X X X X X X X X X X X X

ADD $= \frac{8+9}{2} = 8.5 = \text{MEDIAN}$

$$\boxed{3} \text{ MEAN : AVERAGE} = \frac{\text{Sum}}{\text{No}}$$

$$\frac{5 + 6 + 6 + 7 + 7 + 8 + 8 + 8 + 9 + 10 + 10 + 10 + 10 + 10 + 10 + 11}{16}$$

$$\text{Mean} = \boxed{8.4375}$$

$$\boxed{4} \text{ RANGE : MAX} - \text{MIN}$$

$$11 - 5 = 6$$

Q. 13, 5, 8, 4, 8, 3, 11, 8

$$\boxed{n = 8}$$

$$\boxed{1} \text{ MODE : } 8$$

$$\boxed{2} \text{ MEDIAN : STEP 1: } 3, 4, 5, 8, 8, 8, 11, 13$$

x x x x x x

$$\frac{\text{ADD}}{2} = \frac{8+8}{2} = \boxed{8 = \text{Median}}$$

$$\boxed{3} \text{ MEAN} = \frac{3+4+5+8+8+8+11+13}{8} = \frac{60}{8} = 7.5$$

$$\boxed{4} \text{ RANGE} = 13 - 3 = 10$$

MEDIAN..

FIRST METHOD (CUTTING) CAN ONLY USED FOR RAW DATA.
It cannot be used to find median on Table Data.

So, Lets learn a new method for median to be used on table data later on.

(using method 2)

Q: 13, 4, 12, 8, 6, 2, 10, 12
Find median.

$$n = 8$$

STEP 1: ARRANGE

2, 4, 6, ^{4th}8, ^{5th}10, 12, 12, 13

$$\text{Median} = \left(\frac{n+1}{2}\right)^{\text{th}} \text{term} = \left(\frac{8+1}{2}\right)^{\text{th}} \text{term} = 4.5^{\text{th}} \text{term}$$

4th term 5th term

$$\frac{8 + 10}{2} = 9$$

Median.

NOTE: CUTTING METHOD IS USED FOR RAW DATA
FORMULA IS USED FOR TABLE DATA.

2 TABLE DATA (SINGLE VALUE TABLE)

x/marks	f	$fx = f \times x$
0	20	$20 \times 0 = 0$
1	10	$10 \times 1 = 10$
2	30	$30 \times 2 = 60$
$n = 60$ $\sum f = 60$		$\sum fx = 70$

$0, 0, 0, \dots, 0$ — 20 times
 $1, 1, 1, \dots, 1$ — 10 times
 $2, 2, 2, \dots, 2$ — 30 times

(a) $\text{MODE} = 2$

(b) $\text{MEDIAN} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ term}$

$= \left(\frac{60+1}{2} \right)^{\text{th}} = 30.5^{\text{th}} \text{ term}$

30th term

31st term

$$\frac{1 + 2}{2} = 1.5 = \text{Median}$$

(c) **MEAN:** For Table make a new column = fx

$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{70}{60} = \frac{7}{6}$

Caution: For range we subtract x -values. Do not subtract the frequency values

(d) **RANGE:** $2 - 0 = 2$

③ TABLE DATA (RANGE TABLE)

$a < x \leq b$
↓
class or interval.

$\text{Mid}(x) = \frac{a+b}{2}$	x/marks	 f	$fx = f \times x$
$\frac{5+15}{2} = 10$	$5 < x \leq 15$	20	$20 \times 10 = 200$
$\frac{15+35}{2} = 25$	$15 < x \leq 35$	30	$30 \times 25 = 750$
$\frac{35+40}{2} = 37.5$	$35 < x \leq 40$	10	$10 \times 37.5 = 375$
		$n = 60$ $\sum f = 60$ ↓ Total	$\sum fx = 1325$

(a) MODAL CLASS : (MODE) =

$$15 < x \leq 35$$

(b) MEDIAN INTERVAL.

$$\text{Median} = \left(\frac{n+1}{2} \right)^{\text{th}} \text{ term}$$

$$\left(\frac{60+1}{2} \right)^{\text{th}} \text{ term} = 30.5^{\text{th}} \text{ term}$$

30th term

31st term

Median

$$15 < x \leq 35$$

NOTE: FOR A RANGE TABLE BOTH VALUES WILL BE ON SAME FLOOR.
SO NO NEED TO TAKE AVERAGE.

(c) MEAN: FOR TABLE NEW COLUMNS: 1) fx

2) Middle Value (x)

$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{1325}{60} = \boxed{}$$

$$d \text{ RANGE} = 40 - 5 = 35$$

11 The number of items bought by 10 customers at a local store is shown below.

6 7 5 9 10 7 18 10 7 9

(a) State the mode of this distribution. = 7

(b) Find the median number of items bought.

5, 6, 7, 7, 7, 9, 9, 10, 10, 18
 x x x x x x x x x x

$$\frac{7+9}{2} = \frac{16}{2} = 8 = \text{Median}$$

(b) The table summarises the playing times of each of the 100 tracks on Tom's MP3 player.

Mid (x)	Playing time (t minutes)	Frequency	fx
$\frac{2.5+3.5}{2} = 3$	$2.5 < t \leq 3.5$	5	$5 \times 3 = 15$
$\frac{3.5+4.5}{2} = 4$	$3.5 < t \leq 4.5$	30	$30 \times 4 = 120$
5	$4.5 < t \leq 5.5$	50	$50 \times 5 = 250$
6	$5.5 < t \leq 6.5$	15	$15 \times 6 = 90$

$$\sum f = 100$$

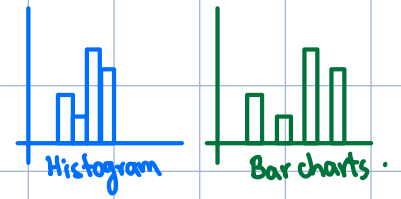
$$\sum fx = 475$$

Calculate an estimate of the mean playing time of the individual tracks.

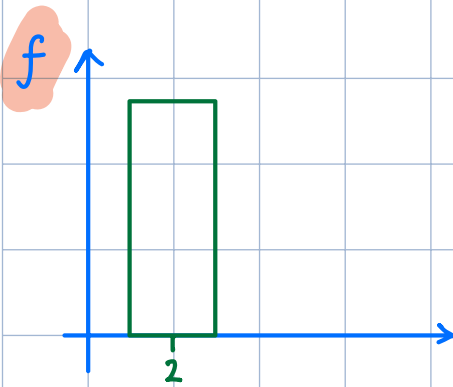
$$\text{Mean} = \frac{\sum fx}{\sum f} = \frac{475}{100} = 4.75$$

Answer (b) minutes [3]

HISTOGRAMS



SINGLE VALUE TABLE



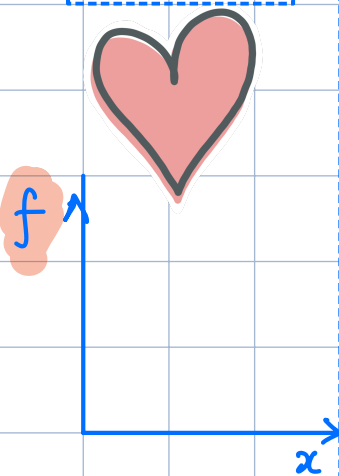
Bar is made in such a way that the number on x-axis is in middle of bar.

RANGE TABLE

New Column = Class width = difference. $(b-a)$

SAME CLASS WIDTH

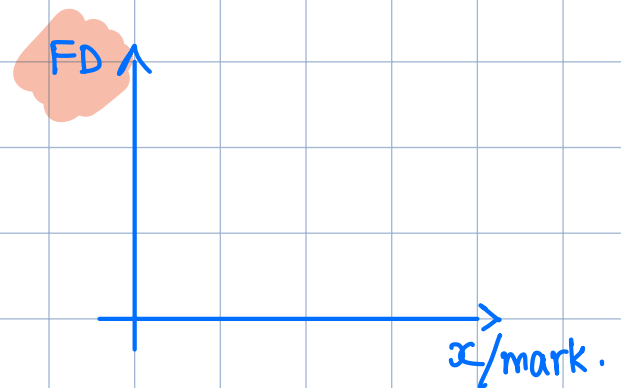
EASIEST HISTOGRAM OF YOUR LIFE



DIFFERENT CLASS WIDTH

New = FREQUENCY DENSITY Column

$$FD = \frac{\text{FREQUENCY}}{\text{CLASS WIDTH}}$$

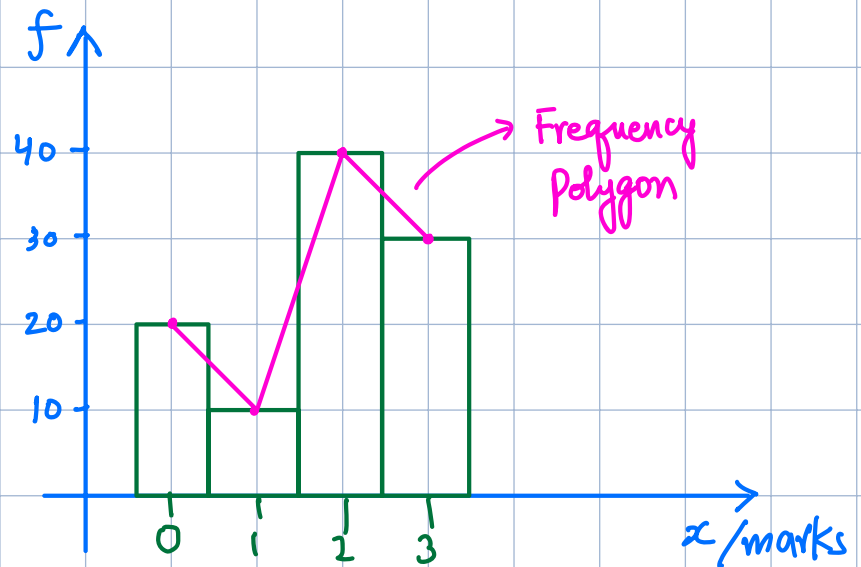


FREQUENCY POLYGON: MAKE A HISTOGRAM AND JOIN MIDPOINTS OF TOP OF EACH BAR.

There is no need to join the lines to x-axis at ends.

1 SINGLE VALUE TABLE .

x /marks	f
0	20
1	10
2	40
3	30



NOTE: In single value table if there is a bar at zero on x-axis, move the whole x-axis one box forward.

2 RANGE TABLE (SAME CLASS WIDTH)

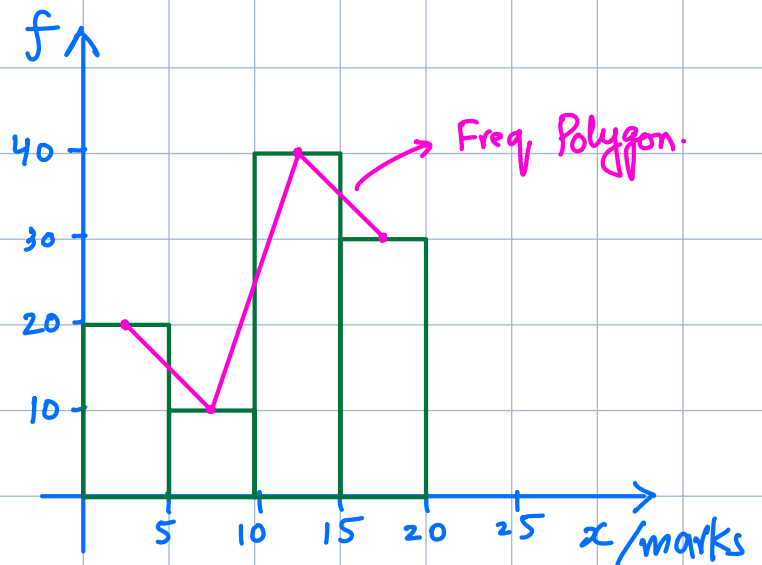
STEP 1: Make new column = class width = $b - a$ (Difference)

	x -axis	y -axis
CLASS WIDTH	x /marks	f
$5 - 0 = 5$	$0 < x \leq 5$	20
$10 - 5 = 5$	$5 < x \leq 10$	10
$15 - 10 = 5$	$10 < x \leq 15$	40
$20 - 15 = 5$	$15 < x \leq 20$	30

↓
If class width is same

we do not need it anymore.

Draw Simplest histogram of your life.



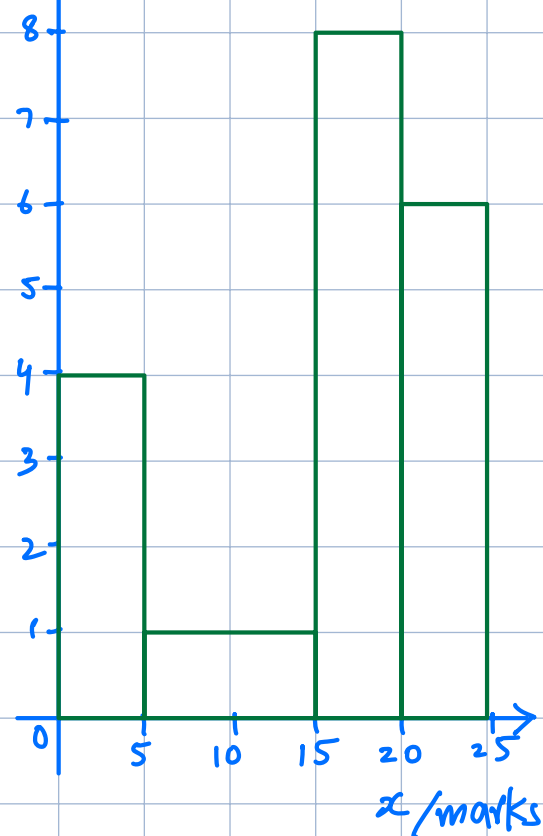
3 RANGE TABLE (DIFFERENT CLASS WIDTH)

x-axis

y-axis

FD

CLASS WIDTH	x/marks	f	FD = $f \div \text{CW}$
$5-0=5$	$0 < x \leq 5$	20	$20 \div 5 = 4$
$15-5=10$	$5 < x \leq 15$	10	$10 \div 10 = 1$
$20-15=5$	$15 < x \leq 20$	40	$40 \div 5 = 8$
$25-20=5$	$20 < x \leq 25$	30	$30 \div 5 = 6$

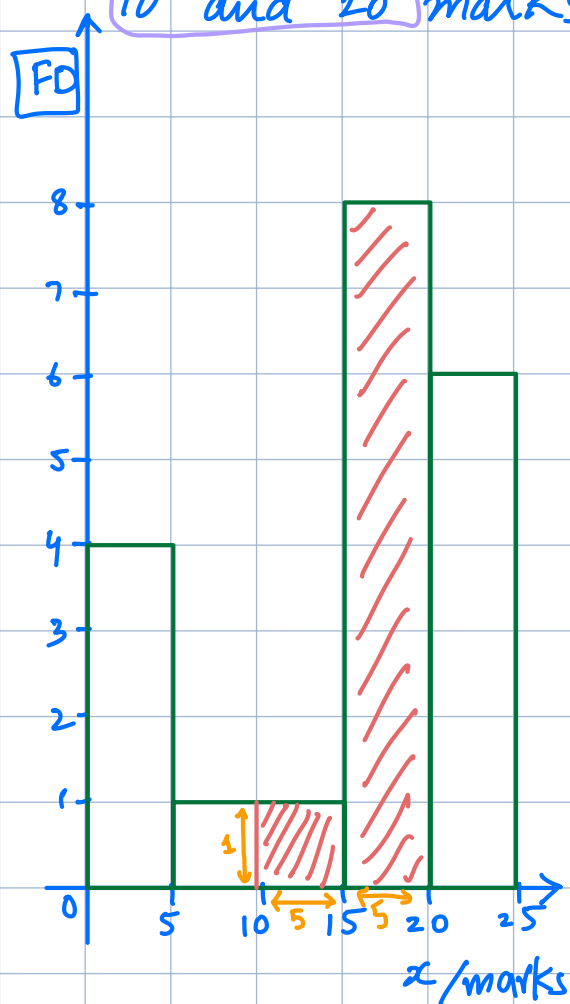


AFTER YOU MAKE A FREQUENCY DENSITY HISTOGRAM.

$$\left[\begin{array}{c} \text{AREA UNDER} \\ \text{GRAPH} \end{array} \right] \text{ OF } \left[\begin{array}{c} \text{FREQUENCY} \\ \text{DENSITY} \\ \text{HISTOGRAM} \end{array} \right] = \left[\begin{array}{c} \text{NUMBER} \\ \text{OF THINGS} \end{array} \right]$$

Area under graph

Q! Find number of students who scored between 10 and 20 marks?



$$\begin{aligned} & (1 \times 5) + (5 \times 8) \\ & 5 + 40 \\ & 45 \text{ students.} \end{aligned}$$

CUMULATIVE FREQUENCY (12 MARKS)

(TOTAL)

One day, garage A records the amount of petrol bought by the first 120 customers.

The results are summarised in the table below.

$$\rightarrow \text{Total} = n = \sum f = 120$$

FREQUENCY TABLE

Petrol (k litres)	$0 < k \leq 10$	$10 < k \leq 20$	$20 < k \leq 30$	$30 < k \leq 40$	$40 < k \leq 50$	$50 < k \leq 60$	$60 < k \leq 70$	$70 < k \leq 80$
Number of customers (f)	9	13	36	30	16	9	5	2

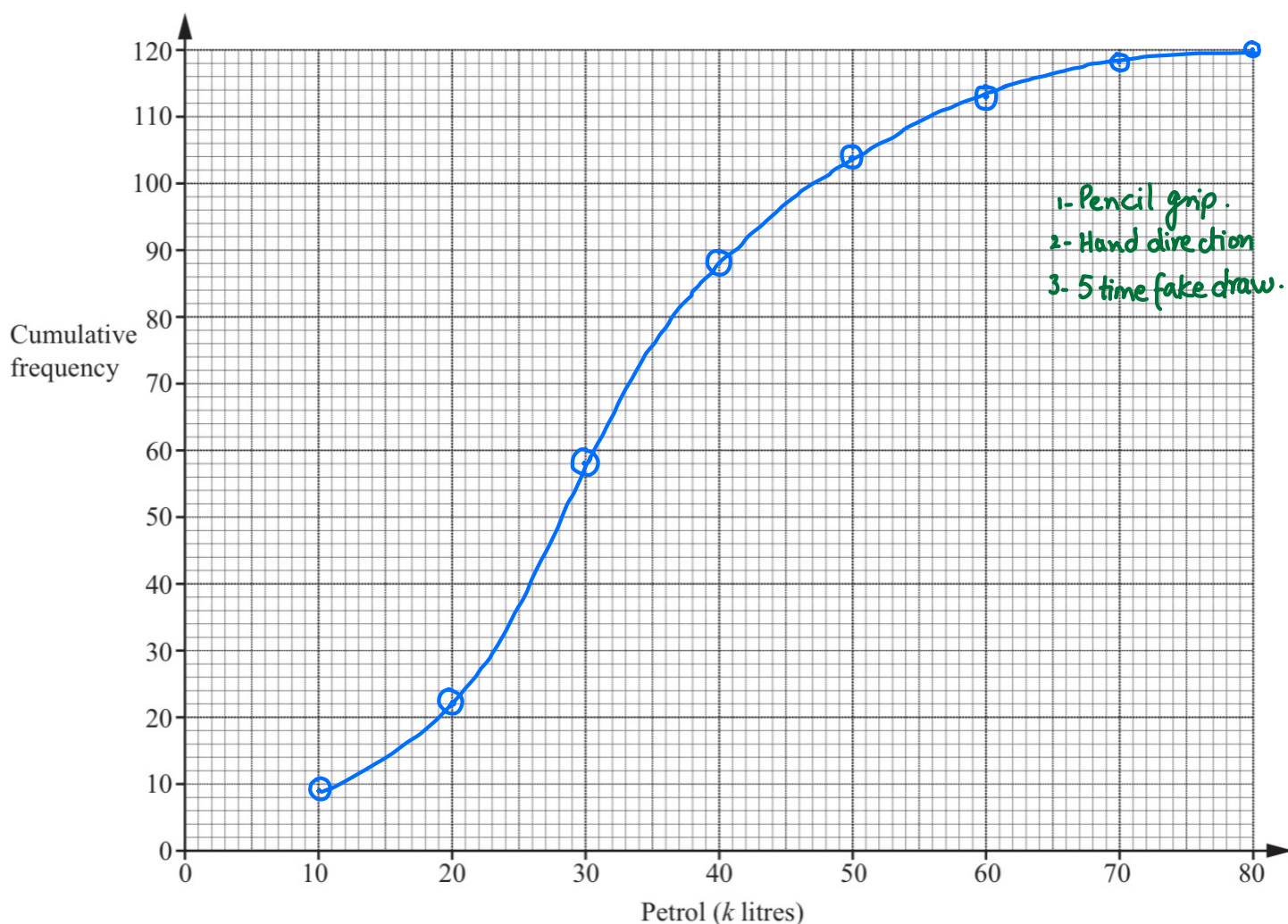
(a) Complete the cumulative frequency table below.

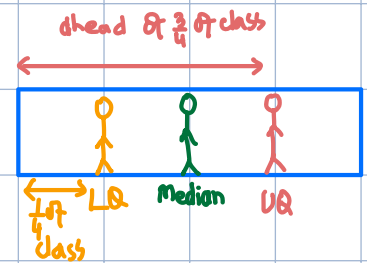
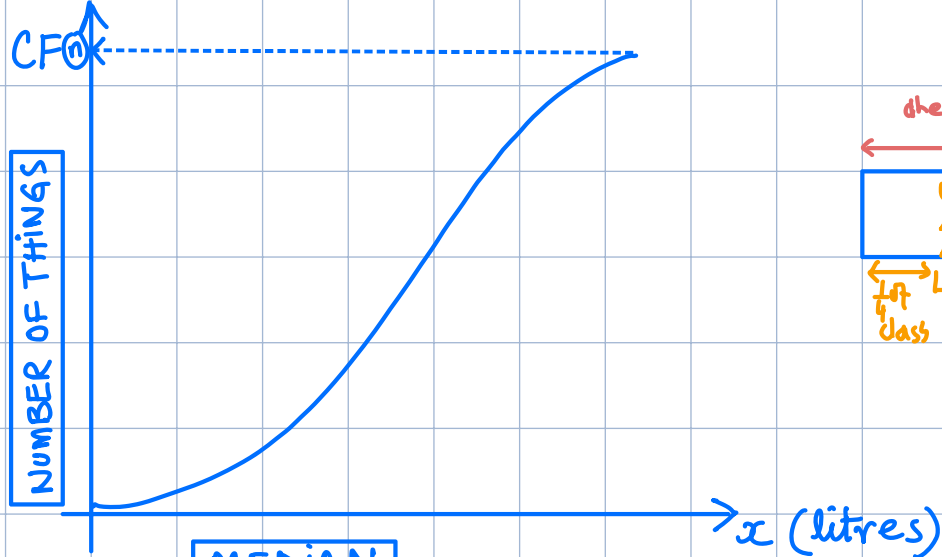
CUMULATIVE FREQUENCY TABLE

Petrol (k litres)	$k \leq 10$	$k \leq 20$	$k \leq 30$	$k \leq 40$	$k \leq 50$	$k \leq 60$	$k \leq 70$	$k \leq 80$
(TOTAL) Cumulative frequency	9	22	58	88	104	113	118	120

(10, 9) (20, 22) (30, 58) (40, 88) (50, 104) (60, 113) (70, 118) (80, 120)

(b) On the grid below, draw a cumulative frequency curve to represent this data.





MEDIAN

UPPER QUARTILE

LOWER QUARTILE

INTER QUARTILE RANGE

PERCENTILE

X-AXIS

Caution Formula

for median is $\left(\frac{n+1}{2}\right)^{\text{th}}$ term every where except cumulative freq. where median = $\frac{n}{2}$

1) **MEDIAN** = $\frac{n}{2} = \frac{120}{2} = \cancel{60}$ check x-axis for Median(M)

Median(M) = 30.5

2) **UPPER QUARTILE** = $\frac{3n}{4} = \frac{3(120)}{4} = \cancel{90}$ check x-axis for UQ

UQ = 40.5

3) **LOWER QUARTILE** = $\frac{n}{4} = \frac{120}{4} = \cancel{30}$ check x-axis for LQ

LQ = 23

4) **INTER QUARTILE RANGE** = upper Quartile - Lower Quartile

$$= 40.5 - 23$$

$$= 17.5$$

Please make sure to subtract x-axis values.

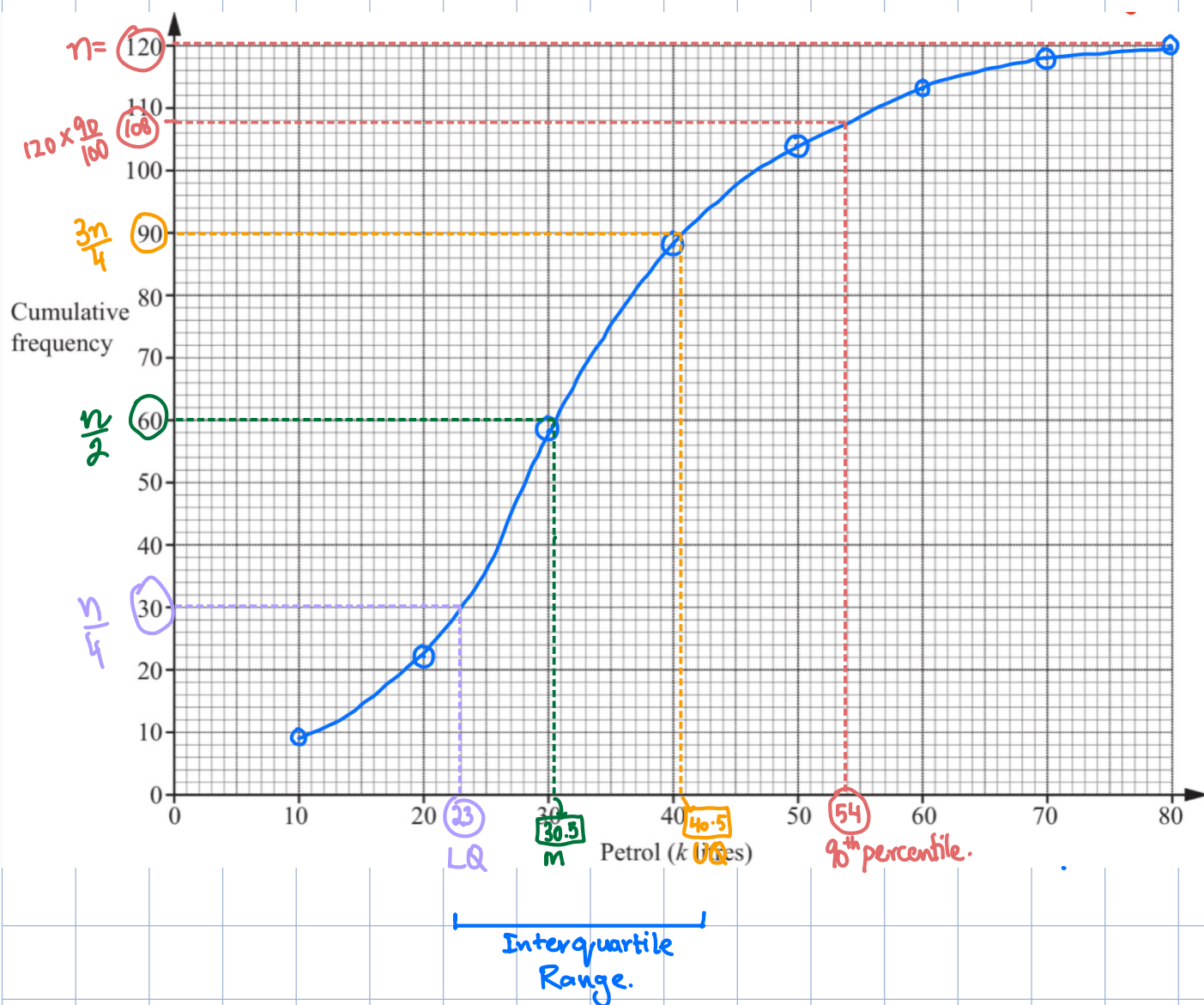
5) **PERCENTILE** (For k^{th} Percentile) = $n \times \frac{k}{100}$ = check x-axis.

For example if you have to find 90^{th} percentile :

$$120 \times \frac{90}{100}$$

= ~~108~~ check x-axis for percentile.

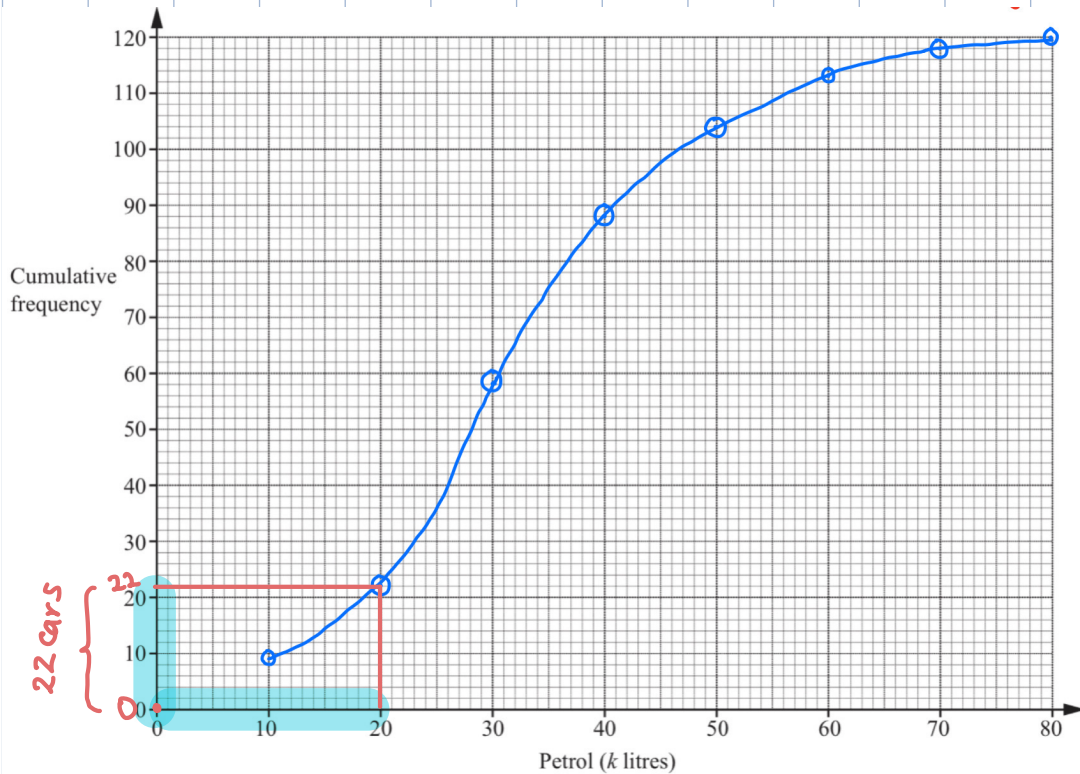
$$90^{\text{th}} \text{ percentile} = 54$$



Y-AXIS : NUMBER OF THINGS.

y-axis

(i) Find number of cars who bought less than 20 litres.

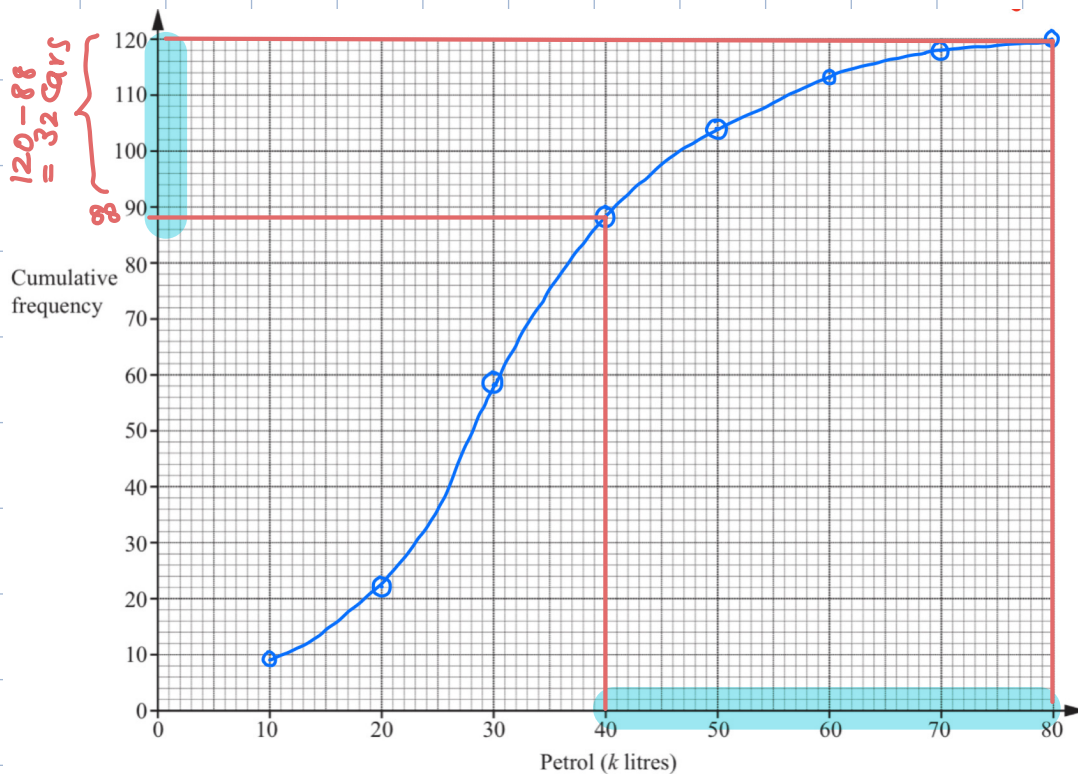


between 0 and 20

$$22 - 0 \\ = 22 \text{ cars}$$

y-axis

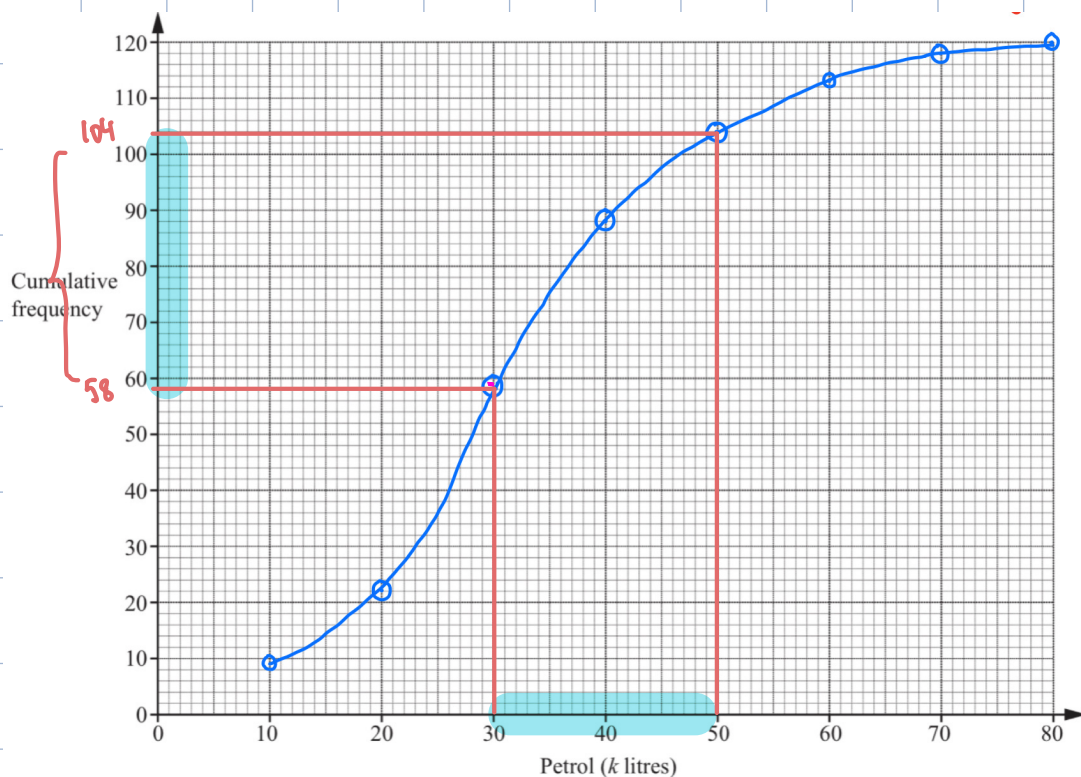
(ii) Find number of cars who bought more than 40 litres.



between 40 and 80

$$120 - 88 \\ 32 \text{ cars.}$$

(iii) Find number of cars who bought between 30 and 50 litres.



between 30 and 50

$$104 - 58$$

46 cars.

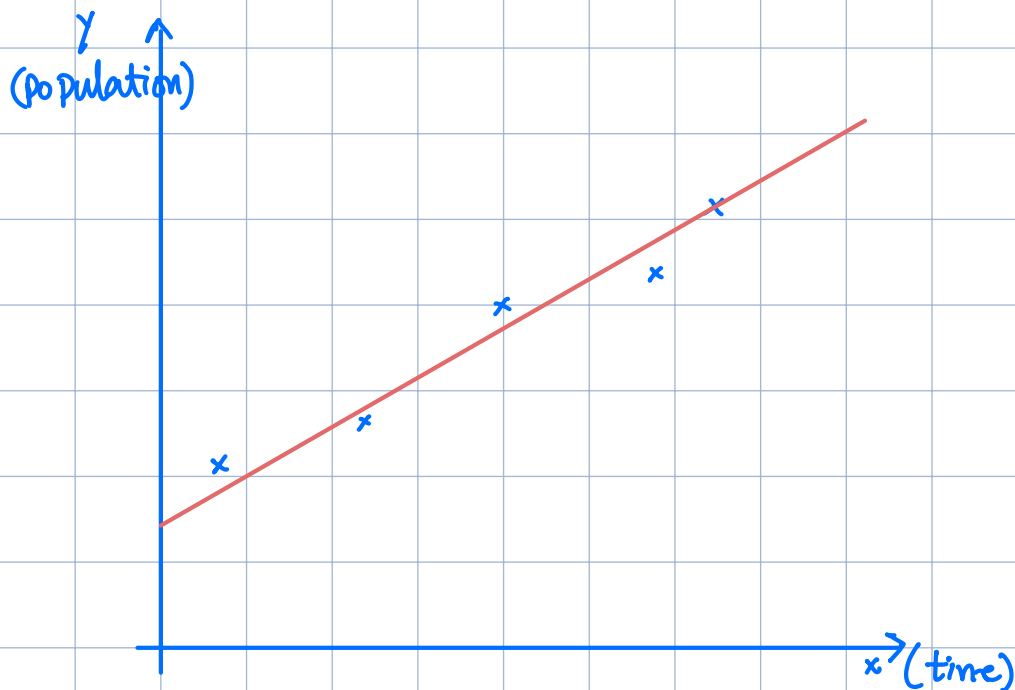
NUMBER OF THINGS

Area under graph
of a Freq-Density
Histogram.

y-axis of
a Cumulative
frequency graph.

SCATTER PLOTS

(LINE OF BEST FIT) (CORRELATION).



Concept

Job of line of best fit is to show trend of plots

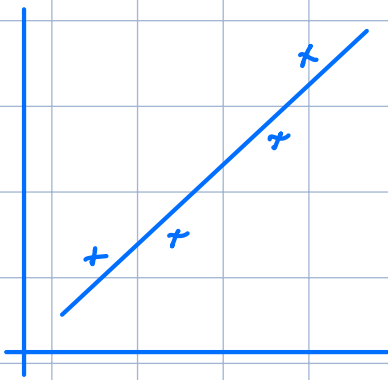
LINE OF BEST FIT:
(TREND LINE)

D'O's

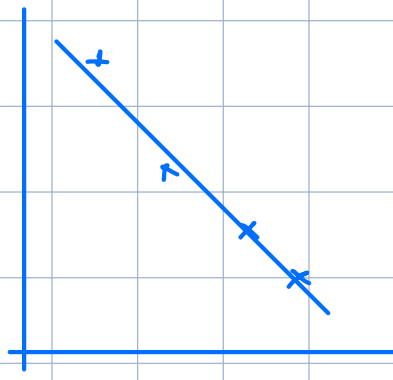
- 1) Average line through the points.
- 2) Almost equal balance of points on each side of line.
- 3) Line of best fit must represent TREND of points

DONT'S

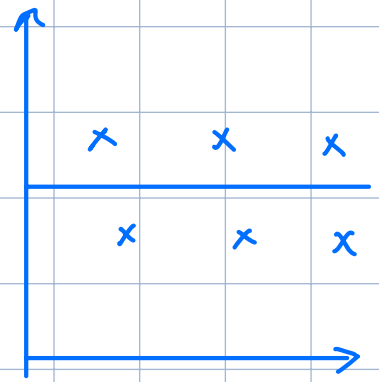
- 1) NEVER FORCE THE LINE THROUGH ORIGIN (0,0)
- 2) NEVER FORCE THE LINE THROUGH FIRST AND LAST POINT.
- 3) NEVER FORCE LINE TO TOUCH MAXIMUM POINTS.



POSITIVE
CORRELATION



NEGATIVE
CORRELATION



NO (Zero)
CORRELATION

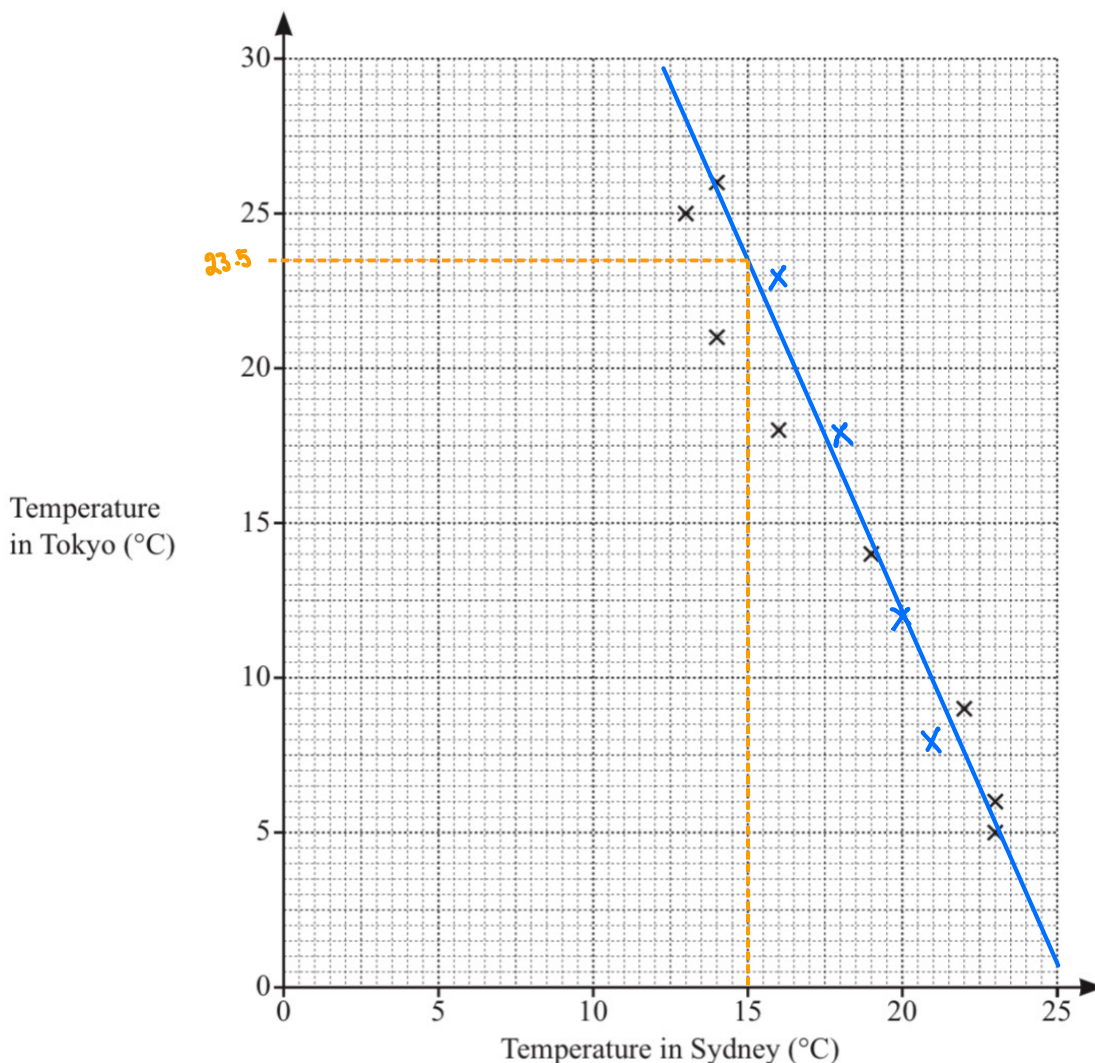
2 The table shows the average monthly temperatures ($^{\circ}\text{C}$) in Tokyo and in Sydney one year.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Temperature in Sydney ($^{\circ}\text{C}$) (x)	23	23	22	19	16	14	13	14	16	18	20	21
Temperature in Tokyo ($^{\circ}\text{C}$) (y)	5	6	9	14	18	21	25	26	23	18	12	8

(16,23) (18,18) (20,12) (21,8)

- (a) Complete the scatter diagram.
The first eight points have been plotted for you.

[2]



- (b) What type of correlation is shown by the scatter diagram?

Negative [1]

- (c) Draw a line of best fit.

[1]

- (d) The following year, the average temperature in Sydney during May was 15°C .

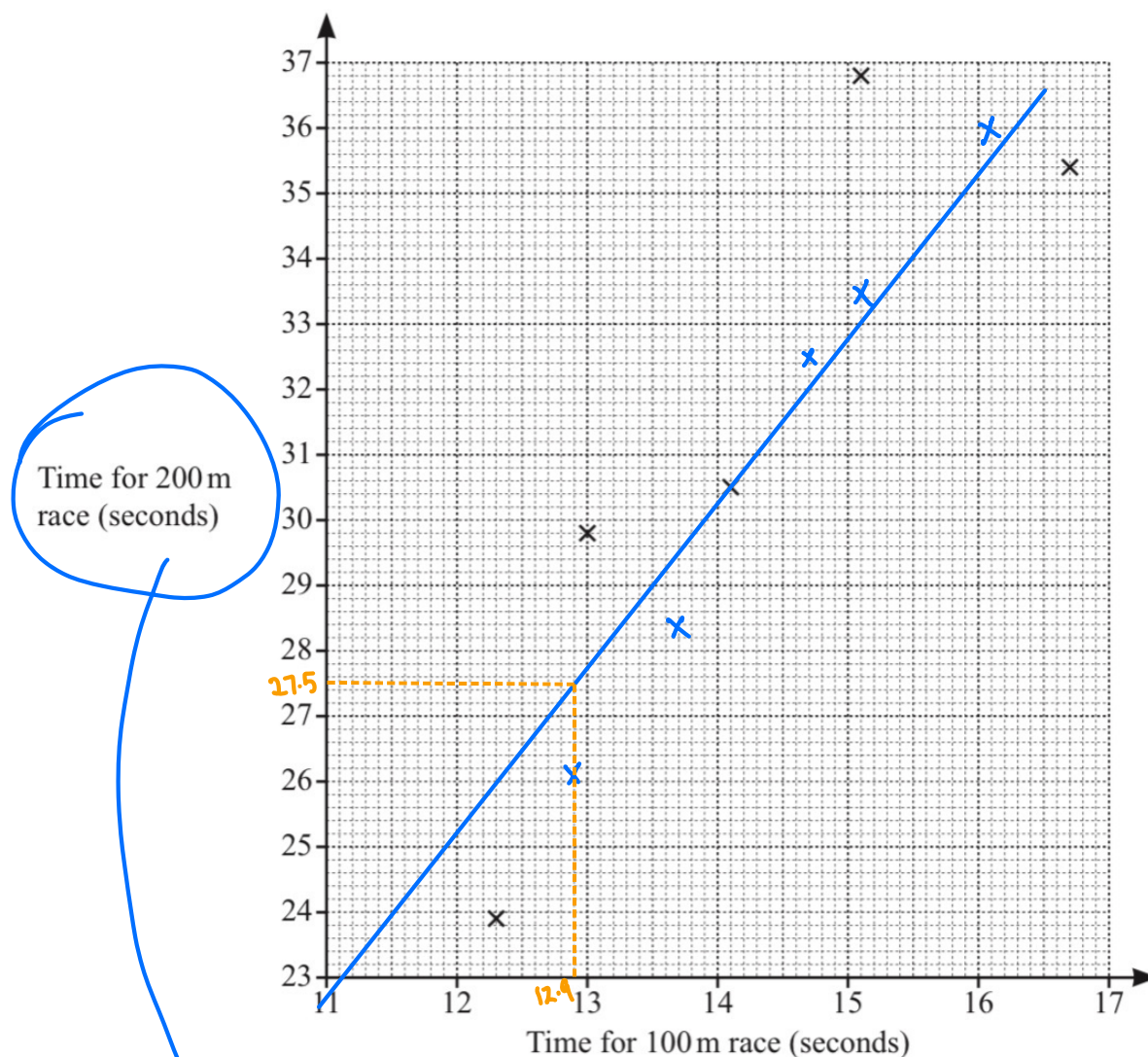
By using your line of best fit, estimate the average temperature in Tokyo that May.

y-axis 23.5 $^{\circ}\text{C}$ [1]

- 2 Ten boys ran in a 100 m race and a 200 m race. The table below shows their times in seconds.

Time for 100 m race (x)	12.3	14.1	15.1	16.7	13.0	14.7	13.7	12.9	15.2	16.1
Time for 200 m race (y)	23.9	30.5	36.8	35.4	29.8	32.5	28.4	26.1	33.5	36.0

- (a) Complete the scatter diagram.
The first five points have been plotted for you.



- (b) What type of correlation is shown in the scatter diagram?

Positive

- (c) Draw a line of best fit.

y-axis = 27.5

- (d) Another boy recorded a time of 27.5 s in the 200 m race.

x-axis

Use your graph to estimate the time it would take him to run 100 m.

12.9

..... s [1]